DOCUMENT RESUME

ED 377 200 TM 022 339

AUTHOR Nitko, Anthony J.

TITLE A Model for Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students.

PUB DATE Jul 94

NOTE 31p.; Paper presented at the Association for the Study of Educational Evaluation in Southern Africa's (ASEESA) International Conference on Educational Evaluation and Assessment (2nd, Pretoria, South Africa, July 1994).

PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Criterion Referenced Tests; *Curriculum Based Assessment; Educational Assessment; Educational Policy; Elementary Secondary Education; Foreign Countries; *Licensing Examinations (Professions); Models; National Competency Tests; *Norm Referenced Tests; School Restructuring; *Selection; Student Evaluation; *Test Construction

IDENTIFIERS *Botswana; High Stakes Tests; National Curriculum (Botswana)

ABSTRACT

This paper evaluates existing national examination development processes in light of changes created by curriculum reform, and the restructuring and expansion of basic education. A model is proposed that creates a strong alignment of national examinations and a national basic education curriculum. Practical examples from Botswana, where the model is being implemented, make the discussion concrete. In a high-stakes environment, articulating curricula and examinations requires at least three components: (1) a formal policy statement about the need for this articulation; (2) adoption of a curriculum-driven examination development model that details the steps in developing examinations; and (3) establishment of an oversight committee to ensure satisfactory implementation. A curriculum-driven examination development model differs from a syllabus or gazette-driven model in requiring that the national curriculum be the center of the examination development process. Since curriculum defines content and performance levels, the examination committee and chief examiner play more facilitative roles than have been traditional. Expected benefits of such a model are outlined. (Contains 11 references, 4 tables, and 4 figures.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ANTHONY J. NITKO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 377 200

A Model for Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations For Certification and Selection of Students

by

Anthony J. Nitko
University of Pittsburgh

July 1994

A paper presented at the Association for the Study of Educational Evaluation in Southern Africa's (ASEESA) Second International Conference on Educational Evaluation and Assessment, Pretoria, South Africa, July 1994

TM022339

BEST COPY AVAILABLE

ABSTRACT

A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students

by
Anthony J. Nitko
University of Pittsburgh

This paper evaluates existing national examination development processes in light of changes created by curriculum reform, and restructuring and expanding basic education. A model is proposed which creates a strong alignment of national examinations and a national basic education curriculum. This paper uses practical examples taken from the Botswana context where the model is in the process of being implemented.

In a "high stakes" environment, where examinations determine who is certified and selected for further education, examination development cannot proceed independently from national curriculum reform. It is necessary for persons at all levels of the educational enterprise to understand that teaching the new curricula in all their important nuances is identical to preparing students for the national examinations. Articulating curricula and examinations requires at least three components: (1) a formal policy statement about the need for this articulation, (2) adoption of a curriculum-driven examination development model which gives the details of the specifics steps required for developing examinations, and (3) establishment of an oversight committee to assure that the required policy and development model are implemented to the satisfaction of the Ministry of Education. The model presented delineates the specific steps and technical procedures which examination developers should follow to assure curricula and examinations are fully aligned and fair.

A curriculum-driven examination development model is different than a "syllabus" or "gazette" driven model. A curriculum-driven model requires that the national curriculum be the center of the examination development process and the decisions about what and how to examine are heavily influenced by the curriculum's stated learning outcomes. A curriculum-driven approach also implies quite a different role for each subject area's "examination committee" and "chief examiner". Since the curriculum defines the content and performance levels to be examined, the examination committee and chief examiner play more facilitative roles than have been their traditional bodies.

Among the expected benefits from implementing curriculum-based examinations are improved: (1) curriculum implementation, (2) examination fairness, (3) assessment of national educational progress, (4) curriculum evaluations, (5) career and job guidance, (6) teacher attention to areas of needed instruction, (7) in-service training, and (8) improved continuous assessment.

A Model for Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students

by

Anthony J. Nitko
University of Pittsburgh[1]

Introduction

For many countries the period following independence, democratization, or political change is a time of rapid educational change: Universal education begins to expand, new school facilities are built, new curricula are developed, and new instructional methods are devised. As such changes begin to take hold, there arises the need to consider whether the existing examinations are still appropriate and serve the best interest of the nation.

Oftentimes curriculum and school-based reforms result in poor congruence between what is intended by the curricular innovations and what appears on the national examinations. As this lack of congruence grows, the examinations may interfere with educational reform especially if the examinations are used for selection. The "high stakes" nature of certification and selection examinations make them powerful forces in shaping what teachers do in the classroom. Unless examinations are properly aligned with curriculum reforms and desired pedagogical practices, it is extremely difficult to implement changes as rapidly as policy makers wish.

In this context of educational reform and national selection examinations, some nations find growing dissatisfaction and criticisms of the examinations. Among the criticisms frequently expressed are the following (Nitko, 1989):

(1) Test results appear to be insensitive to improvements in educational inputs and to teachers' and parents' perceptions of pupils accomplishments.

(2) Test reports do not describe the knowledge, skills, and abilities which students have learned. As result, policy makers and curriculum developers do not know what areas of the curriculum to improve.

---

[1]Correspondence concerning this paper may be sent to Professor Anthony J. Nitko, University of Pittsburgh, 5B26 Forbes Quadrangle, Pittsburgh, PA 15260, USA. Internet:ajnitko+@Pitt.edu.

(3) Examination results provide a poor basis for advising students for vocational and career development.

(4) The correspondence between the learning objectives stated in the official curriculum and the questions which appear on any one year's examination is often unclear for teachers. The result is that teachers stop teaching the official curriculum and use past examination papers as the teaching materials.

(5) Educators at all levels find it disconcerting that at certain points in a student's schooling levels there is reliance on using a single "high stakes" examination result which ignores many years of student performance in the classroom.

(6) The breadth and richness of new curriculum reforms are ignored by teachers who take it upon themselves to narrow the curriculum to those tasks likely to appear on the examination.

Until there are sufficient resources to assure places in higher levels of schooling for all students, there will be a need to select students. However, this selection need can be serviced in a way that permits criticisms of examinations to be addressed. This can be done by developing curriculum-driven examinations that possess criterion-referenced qualities, but which do not lose their norm-referencing ability. The purpose of this paper is to describe a model for developing such examinations.

Norm-Referencing and Criterion-Referencing

Before discussing the model, I would like to clarify the nature of criterion-referencing and norm-referencing as these concepts apply to national examinations. ***Norm-referencing*** refers to interpreting a student's test score by comparing it to the scores of other students in a population. The population against which the student is compared is called the ***norm-group***. ***Criterion-referencing*** refers to interpreting a student's score by comparing it to a domain of performances that the student is expected to learn as a result of instruction in a given curriculum. The domain of curriculum objectives or learning targets is called the ***criterion***.

The referencing of students' raw marks is necessary for all examinations because the raw marks cannot be properly interpreted without referencing. For example, if you know that a student has obtained 68 marks on an examination, that information alone does not describe the student's performance. However, you could reference this score to the population who took this examination. You might find, for example, that the student's marks were higher than 85 percent of the population. Thus, you could make the interpretation that this student performed quite-well -- relative to other students. The norm-referencing, however, provides you with only an incomplete interpretation of the student's performance. Criterion-referencing rounds out the picture. For example, consider once more your hypothetical student who received 68 marks. Perhaps a mark of 68 means that this student mastered only 50 percent of the performances expected by the curriculum objectives. Thus, even though this student outperformed 85 percent of the population, the student's absolute level of achievement leaves much to be desired.

You should note that both kinds of referencing are desirable in order to interpret an individual's scores validly. Criterion-referencing and norm-referencing are not mutually exclusive referencing schemes. Rather, they are complementary schemes; they are obverses of the same coin.

However, it is possible to obtain both kinds of referencing from a single test ***only if special procedures are followed when designing and producing the test***. Valid norm-referencing is possible, for example, only when the norm-group against which a student's score is compared consists of the entire population of similar students ***or*** when we have followed special procedures to obtain a representative sample from the population. Similarly, valid criterion-referencing is possible only when we have assessed a student on the entire domain of curriculum learning targets or when we have followed special procedures to obtain a representative sample of learning targets from the domain of targets specified in the curriculum. If the special procedures are not followed, one or the other type of referencing will be weakened and, thus, less valid. This paper focuses on the procedures that should be followed for developing curriculum-based examinations so that criterion-referenced interpretations may be made. I turn to that process in the next section.

A Model for Curriculum-Driven Criterion-Referenced Examination Development

Figure 1 shows a process model for developing curriculum-driven examinations. The model shows the major stages of examination development in terms of what outcomes are expected at each stage of the process. The stages begin at the lower left of Figure 1, and move to the right. There are nine major stages in the process. In the next sections of this paper, I discuss these steps in more detail. However, before doing this I will briefly describe what I mean by curriculum.

Insert Figure 1 here

What Is Curriculum?

A major feature of the model shown in Figure 1 is that it depicts the process of examination development as beginning and ending with the curriculum. Thus, in order to understand and to implement the model, we need to come to some understanding of what curriculum is. At first thought, it might seem easy to define the curriculum for which assessment is to be planned. This is far from reality, however. The first problem is that there is no standard concept of what constitutes a curriculum. One or more of the following is often considered to be "the curriculum" (Posner, 1992):

- the scope and sequence
- the syllabus
- the content outline
- the textbooks and teacher's guides
- the planned experiences for the students

The second problem is that there may be five curricula operating in the schools at the same time. In theory, one or more of these may be used for examination development. These five are (Posner, 1992):

- the official curriculum -- that is, what is found in official statements and materials.
- the operational curriculum -- that is, what the teachers actually deliver to the students and for which they hold students accountable through their own assessments.
- the hidden curriculum -- that is, what the students actually understand and experience through being in school, including what is taught about norms, values, roles, authority, legitimacy of certain knowledge, and so on.
- the null curriculum -- that is, what is not taught and why it is not taught.
- the extra curriculum -- that is, the planned experiences outside of the school subjects in which students learn such things as fair play, competition, leadership, and how school subjects are valued in relation to sports and other nonacademic activities

In my view, ***curriculum*** is both a means and a rationale through which schools can coordinate educational experiences, materials, and teaching. These, in turn, guide schools in creating the conditions in which students can learn. A properly developed curriculum includes more than statements of goals, standards, and learning targets. It must also provide full educational, social, and moral rationalizations, of not just educational outcomes, but also the educational process through which students should progress. Assessment tasks (examination questions), even those that are well-constructed, authentic, interesting, performance-based, and motivating, cannot be used on their own to fully rationalize the desired goals, processes, and outcomes of the educational enterprise.

A curriculum's rationale should present a compelling justification of the full range of a student's educational experience in a subject area. This includes rationalizing the content

teachers should cover, the educational outcomes students should attain, the scope and sequences teachers should follow, and the educational activities that give students opportunities to reach the desired learning outcomes. This justification comes about by weaving together ***many*** ideas, not just those of the discipline(s) underlying the subject-matter. A curriculum must also explain such factors as the moral and social philosophy that justifies school experiences, the pedagogy that sets the conditions for learning, and the theories and empirical findings from various areas of educational and social research. Included in educational research are the fields of human learning and cognitive psychology. When a curriculum is fully designed and satisfactorily implemented, it becomes the foundation on which schools can build both instruction and assessment.

Stage One. Define the Achievement Outcome Domain Intended by the Curriculum

Begin and end with curriculum Returning now to Figure 1, we notice that the model shows that the examination development process begins with the curriculum. This follows from the fact that if curriculum is to rationalize the educational process, it must also be the rational basis for educational assessment. Assessing the important outcomes intended by the curriculum becomes the major focus of Stage One and of all the remaining stages of the examination development process. Therefore, the first requirement, which is represented by Stage One, is that the curriculum should be the master of both the educational and examination enterprises (cf. Madaus, 1991).

Harnessing high-stakes forces In the presence of high-stakes examinations, authorities tend to judge the quality of teachers and headmasters, at least in part, by their students' examination performance. This has a significant impact on the operational curriculum and creates a discrepancy between the operational and official curricula. It also creates a force in the educational system: A force that motivates teachers to teach to the examination, while de-emphasizing or not teaching those objectives in the official curriculum which they believe will not be on the examination. In the presence of high-stakes examinations, the key to making the operational curriculum correspond more closely to the official curriculum is not to try to eliminate this force. Rather it is to create examinations that are closely aligned or driven by the official curriculum. As a result, the force that motivates teachers to teach to the examination is harnessed and directed to the desired end: ***Teaching to the examination is essentially teaching the official curriculum***. Stage One is the first step to accomplishing this curriculum-to-examination alignment.

Validity evidence begins with Stage One From a psychometric perspective, the validity of any curriculum-driven criterion-referenced assessment depends in large part on how well the curriculum learning targets are defined and how faithfully the assessment tasks represent these important learning outcomes. This means that examination developers at every stage of the process, must constantly judge the quality of the tasks they develop against their faithfulness to the intended outcomes of the curriculum. Positive and negative evidence for the validity of

curriculum-driven assessments begin to accrue when the examination tasks are ***initially*** conceptualized, beginning with Stage One. Validity evidence continues to accumulate through each stage of the development process, since you verify the curriculum-based integrity of the tasks at every stage.

Multiple sources define the curriculum Operationally, Stage One requires examination developers to work closely with curriculum developers to fully define the curriculum domain on which the examination will be developed. In practice there is no single document that satisfactorily defines a curriculum in all its important nuances. As a practical matter, therefore, Stage One requires reviewing and synthesizing a variety of sources including the curriculum developer's ideas, cognitive theory, curriculum theory, the content syllabus, the curriculum goals, classroom activities of the best teachers, and instruction materials, such as textbooks and practice materials.

Stage Two. Analyze the Curriculum

Obtaining some sense of what is the curriculum is only the first step in the assessment development process. The next step is to clearly identify and organize the intended learning outcomes of the curriculum so that an assessment system and plan can be created. This is the activity shown as Stage Two in Figure 1. Before an assessment system and plan can be developed for a curriculum area, it is necessary to make clear the meaning of the curriculum. That is, you need to identify the assumptions the curriculum makes, the goals and outcomes that are specified, the correspondence of these specified outcomes to a framework that organizes the goals and outcomes, and the priorities among all the competing outcomes and components of the curriculum. The basic output of this analysis is a document that is a well-organized specification of the cognitive and noncognitive outcomes which you should assess in one form or another. The outcomes need not be specified as narrow behavioral objectives. However, the students' learning targets should be clear.

Mapping the curriculum There are at least two benefits that come from this curriculum analysis. One benefit is the production of a kind of "curriculum map" that further clarifies (a) those parts of the curriculum on which students should be formally assessed and (b) who should do the assessing. Some parts of the curriculum will be better assessed at the local school level by teachers. Other parts may be better assessed by examinations external to the classroom. These external assessments may include assessments developed by regional panels of teachers. Others may be external assessments set at the national level.

An important point is that when you review the curriculum analysis, it will be apparent that curriculum-driven assessment must include a formal mechanism for teacher-based continuous assessments. Curriculum-driven assessment must not be limited to less frequent end-of-year, to end-of-term, or to national examinations. Further, the curriculum analysis stage will make it clear, too, that there must be a logical consistency to what the assessments tasks require of the students at all levels of the educational experience, from the classroom to the articulated

standards in the curriculum.

Seamlessness I refer to this consistency as seamlessness (Nitko, 1994b). **Seamlessness** means that, regardless of whether the assessments are produced by teachers or by a national examining board, they are immediately recognized by school officials, teachers, students, parents, and the public as requiring the same complexities of knowledge, processes, skills, and abilities, that are not only desirable to learn, but which in fact have been taught in the school over a considerable period of time. In this way, teaching and assessment become aligned and integrated.

Seamlessness is desirable in either high or low stakes assessment, but it is especially necessary if student accountability is associated with the examination. This is the case for certification and selection decisions. In high stakes situations, the assessments will drive the teaching. Out of moral necessity, teachers must orient their teaching to maximize the students' chances of meeting the high stakes standards. If assessments are not seamless and fully representative of the curriculum, the teachers will (and should) ignore those parts of the curriculum that will not count toward the certification or selection decision. If seamlessness is not present, assessments are not properly aligned with the curriculum. There is a tear in the educational fabric and curriculum reform will most likely not be properly implemented.

Curriculum revision A second benefit of the Stage Two curriculum analysis is to help curriculum developers understand the existing curriculum. Assessment-oriented curriculum analysis focuses on the performance outcomes expected of all students. The benefit to curriculum developers lies in the increased insight they obtain into the curriculum's organization, into what requirements various assessments will impose, into the benefits and limitations the assessments could provide, and perhaps into ways the curriculum may be modified. Experience shows, for example, that seldom do curriculum development officers clearly articulate higher-order thinking goals in a way that can be used for either lesson planning or assessment at the classroom level. A curriculum analysis in the context of designing an assessment system and plan may point out such inadequacies. It may lead to expanded and improved curriculum statements and materials.

Stage Three. Assessment Plan Development

My discussion so far has indicated that aligning or linking curriculum and examinations cannot be accomplished satisfactorily unless the curriculum is clearly defined and analyzed. These two steps identify what is to be assessed and who in the educational system will be responsible for developing and administering the various assessments. Once these have been identified, an assessment plan can be fleshed out. This is Stage Three.

Plan for more than an examination An assessment plan should span the full range of school years and not just the certification or selection examination. The plan should describe the assessments expected (a) at the level of continuous, teacher-based assessment, (b) at the level

of the school in the form of termly and/or annual examinations, and (c) as part of the leaving examination. An analysis of the curriculum may be used, for example, to identify the important learning outcomes that should be assessed at each standard or grade every year. Ideally, the curriculum analysis will identify progressions of outcomes that span the school years. If these outcomes are learned by students as they progress through school, the students will most likely be successful on the leaving examination. I have described a curriculum-driven criterion-referenced framework for continuous assessment elsewhere (Nitko, 1994a).

Sampling plan needed The assessment plan will most likely require specifying how to ***sample*** learning outcomes because there are many more outcomes than can possibly be assessed at one time. Every assessment procedure ultimately leads to a narrowing of the operational curriculum to performances that will appear on (Lindquist, 1951). Creating a sampling plan and making it public may help to minimize this narrowing effect. The type of plan I have in mind makes clear to teachers (a) that all important aspects of the curriculum are fair game for the examination, (b) the procedure that the examining body will use to select the assessment tasks, and (c) the weight each part of the examination will have. For example, the plan would make clear to teachers which parts of curriculum will ***always*** be assessed on the examination and which parts will be included ***only*** on a random sampling basis. If teachers believe there is a chance that a part of the curriculum will be assessed each year, they may not be inclined to focus teaching specifically on what has appeared on one or two past examination papers.

Prototype assessment tasks The assessment plan should also include developing prototype assessment tasks and procedures that would be used at various levels of the educational enterprise, that is, used in school-based assessments and in the national examinations. The prototype assessment tasks must be carefully designed to assure that they faithfully represent the important variations and richness of the curriculum. These prototype tasks should include both paper-and-pencil questions as well as performance tasks, open-ended tasks as well as focused single-correct-answer tasks. Alternative assessments should be include where appropriate to assure the intended curriculum outcomes are assessed properly.

Making plans and prototypes public Both the assessment plan and the prototype tasks which operationalize the plan should be made public so that teachers are aware of the level of performance expected of the students. This type of openness is desired in curriculum-driven examination development. It makes clear to educators and the public the curriculum-to-assessment linkages. More importantly, it demonstrates that teaching the curriculum must be taken seriously because doing so will lead to successful performance on the certification and selection examination.

Establishing examination committees It is probably necessary for implementing Stage Three and subsequent stages, to create committees to oversee and monitor examination development. Checks, balances, and intellectual inputs are as important for curriculum-driven criterion-referenced assessment systems as they are for norm-referenced assessment systems. Traditionally, these committees are known as "examination committees" and are often headed by a "chief examiner". However, with curriculum-driven criterion-referencing, it is necessary

for this examination committee to play a somewhat different role. Table 1 shows a comparison of the functions of the traditional syllabus-driven examination committee with the functions of a curriculum-driven criterion-referenced examination committee. The latter committee plays a more facilitative and evaluative role than the former. In the latter case, curriculum is given a more central role in determining the examination tasks and teams of curriculum officers and examination officers work cooperatively in professionally responsible ways. The numbers assigned to functions in Table 1 should not be interpreted literally as steps in a sequence. Rather, they are simply to distinguish the functions. The numerical order in the table is a rough indicator of sequence, however. You can see from the table that the examination committees play central roles at every stage of the test development process.

Insert Table 1 here

In order to maintain consistent high quality examinations across all subject examinations, it may also be necessary to create an oversight committee to coordinate the separate examination committees. Such an oversight committee would assure that curriculum-driven criterion-referenced examination policy is consistently implemented in each curriculum area. It may also recommend policy changes to the Ministry of Education as such needs arise.

Stage Four. Developing Assessment Task Specification

After the assessment plan is created, the next stages become more technical. The fourth stage is one of refining the prototypes tasks so they are valid assessment tools. This refinement is especially important for that part of the plan that applies to the leaving examination. The goal at this stage is to describe the nature of the assessment tasks in sufficient detail that it is clear to those who produce the examinations, which tasks are validly included and which are not. Using task specifications as a basis for setting examination questions increases the consistency of the examinations from year-to-year. This consistency, in turn, increases year-to-year comparability of examination marks. High comparability means that the examinations are fairer to students.

Creating task or item specifications for curriculum-driven leaving examinations is an extra step that has typically not been used in many countries. Figure 2 shows a narrow and highly detailed item specification. This specification would be created by the examining board staff and would be approved by an examination committee. It would then be used by committees of teachers, under the guidance of examination officers, to create individual examination questions. Many such item specifications would need to be created.

Insert Figure 2 here

Experience with these narrow item specifications indicates that they (a) may be too restrictive to examination officers and (b) may produce examination questions that are too stereotyped (Popham, 1992). A broader type of task specification may be more useful. An example of this broader specification is shown in Figure 3.

Insert Figure 3 here

The important point about using task specifications is that they provide a way to guide assessment tasks setters to produce tasks that are consistently faithful to the curriculum outcomes they should be assessing.

Stage Five. Producing and Validating Tasks

Stage Five is the task-setting stage. Unlike typical task-setting exercises, curriculum-driven task-setting has at least three quality assurance procedures to assure that the assessment tasks produced faithfully represent the curriculum. First, the curriculum-driven task specifications are used as guidance for creating the assessment tasks. You will recall that these specifications were created and reviewed in Stage Four to assure that they match the learning targets specified in the curriculum. Second, each task that is set is subject to formal review by a panel of curriculum experts to assure that it matches its respective task specification and that it faithfully assesses the desired curriculum learning target. Third, there should be empirical trialing of the assessment tasks and scoring rubrics to assure that they function properly and that the students interpret them in the way intended. It is essential, too, that scoring rubrics for open-ended and performance tasks be refined using actual responses of students. The empirical steps are more difficult to accomplish for secure tests, such as leaving examinations. Nevertheless, empirical trialing is very important and some accommodation to it is necessary. Some suggestions for these accommodations in secure test situations include (a) trialing tasks two or three years before they are needed, (b) using small samples of students so security can be maintained, and (c) building a large item-bank of trailed tasks from which one may draw samples in any one year.

Stage Six. Assembling the Examinations

The sixth stage is a well-known one: Putting the examination together and producing it. The important point here is that the published version of the examination must be a representative sample of the important learning outcomes in the curriculum and the tasks on it must be clearly recognized as curriculum-based. The operative term here is "representative". The examination "re-represents" the learning domain defined by the curriculum. That is, through appropriate specification and sampling, the examination clearly presents the curriculum outcomes in appropriate proportion and weighting.

Stage Seven. Setting Standards

After producing the examinations, it is necessary to set standards for making decisions such as awarding certificates or selecting students. The processes for standard setting must be carried out very carefully in order to assure they are fair to all and that they represent comparable performance from one year to the next. Although this is an extremely important stage in the process of examination development and use, space does not permit a detailed discussion in this paper. Some reviews of procedures are found in the literature and these should be consulted (e.g., Jaeger, 1989; Livingston and Zieky, 1982). It is important to recognize, however, that standard setting should involve both judgments of well-qualified teachers and educators, and empirical data pertaining to how well candidates perform on the examination tasks.

Stage Eight. Primary Analysis

After administering an examination, the results must be analyzed and reported. Usual procedure are followed to mark or score students' responses, to equate the current year's results with previous years', and to report the scores on a suitable score scale. Excellent reviews of these procedures are given in the literature (e.g., Angoff, 1971; Peterson, Kolen, and Hoover, 1989; Holland and Rubin, 1982).

Many national examination programs stop empirical analyses of the data after the students' results are reported to individuals and government. However, it is most important that examining bodies analyze the quality of the examination. This is necessary even if the examination itself is to be "released" to the public and a new form created the following year.

Examining bodies are mandated to develop high quality assessment products. Criterion-referenced testing technology should be used to improve and maintain the quality of an examination from year-to-year. In other words, a quality control program should be maintained and implement as part of Stage Eight.

If high quality curriculum-driven criterion-referenced examinations are to be created, quality standards need to be specified and adopted as the official policy of the examining body. ***Quality standards*** describe the technical properties that the institution ***requires each examination to meet*** before it may be used to certify and select students. The development steps specified in Stages One through Seven assure that some ***minimal*** quality levels are met, but they do not describe how the institution should ascertain and control the quality of each examination produced. Stating and implementing quality standards guarantees an examination's quality.

A nation's children have a right to expect that their leaving examinations are as valid and technically sound as possible under the practical constraints of cost and limited resources. Instituting quality control monitoring increases examination equity because this monitoring assures that the examinations set this year yield essentially the same results as would have been attained had any other year's examination been used. Quality control monitoring is necessary to assure that each year's assessment is fair to students, is equally representative of the curriculum, and that students are held to comparable standards from year-to-year. For curriculum-driven criterion-referenced tests, quality standards go beyond standards used with norm-referenced tests.

Table 2 shows 21 quality control standards which may be used with curriculum-driven criterion-referenced national examinations. (The list is phrased in terms of objective items, but may be adopted easily to essays and performance tasks.) This list may serve as a starting point for an examining body's policy discussions that lead to a final list of standards the body adopts. Once a list of quality standards is officially adopted, the quality of the examinations the body produces can be measured against the standards and the established criteria. All examinations would be expected to meet all criteria before they are approved for official use.

Insert Table 2 about here

Table 2 groups the standards into three quality control areas: item content, individual items' technical quality, and the quality of the examination marks themselves. Within each area, five or six quality standards are listed. The main point is that the standards list states the qualities which the institution wants every examination to exhibit before it is used officially for student accountability decisions. Each quality standard appearing on such a list must be shown to contribute positively to making an examination highly valid and relevant for the purposes of certifying student competence and selecting students for the next level of schooling.

Listing quality standards is only the first step. In order to implement the standards, they must be measured or otherwise assessed, each by one or more procedures, otherwise the

examining body cannot monitor their implementation. Column three of Table 2 lists one measure for each standard. Other measures could be conceptualized. Further, for each measure, a quantitative criterion is set. The quantitative criterion reflects the minimum level of quality which the institution wants its examinations to meet. These criteria are listed in the rightmost column of Table 2. For example, to measure the relevance and importance of each task that the examination question requires of the examinee, the examining body would construct a simple rating scale which subject-matter experts would use. A panel of experts might include senior teachers, education officers, and university professors. The ratings for each examination question would be averaged and compared to the institutionally established criterion listed in column four. Examination questions that do not meet the criterion would be revised or replaced by others that do meet them.

It could be argued that quality control studies should be part of Stage Five during which test items are developed and trialed. I have chosen to put quality control in Stage Eight, however, for two reasons. First, many national examination programs contain essays, practicals, performance tasks, and other open-ended questions, and, second, the examinations themselves are released to the public once they are administered. In such cases, sufficient empirical data to support quality assurance measurement may not be available at the time the examination tasks are prepared. However, sufficient data are available once the examination is administered nation-wide. Although a post hoc analysis of the quality of an examination is not as desirable as an a priori analysis, it is nevertheless feasible. Monitoring the quality of already administered examinations will provide indicators of their quality and will point to areas for which better assessment development procedures need to be implanted. The point is, you should not avoid measuring examination quality even though you do not do extensive trialing of test tasks before the examination is administered.

It should be noted with regard to quality control, that once the quality standards are officially approved, it is necessary to give one person the responsibility of monitoring their implementation for all examinations the agency produces. A quality control examination officer would report on all examinations which fail to meet the official quality standards and which need to be improved. Past examinations should be analyzed first and their quality described. This procedure would identify those subjects whose examinations have a history of poor quality and which should be targeted first for improvement.

Stage Nine. Secondary Analyses

If the first eight stages of the model are followed, the national examination results will be a rich source of data useful for educational policy analysis and curriculum reform. The processes described in the model assure that the examinations are aligned with the curriculum and that they properly represent (or "re-present") it. Thus, the results of any one examination can be meaningfully broken down by curriculum topic or type of thinking skill required, and reported at the school, region, or national level. The results of the secondary analyses of the

examination data maybe fed back into the examination development system to improve assessments and into the curriculum unit to suggest areas of curriculum that need improvement.

In this regard, it is necessary to identify the stakeholders for receiving the results of secondary analyses so that the reports may be tailored to their needs. Table 3 shows some stakeholders and examples of the types of reports each may find useful to their missions. Table 4 is an example of a headmaster's (principal's) report in one curriculum area.

Insert Tables 3 and 4 about here

Fleshing-Out the Model In the Local Context

The model shown in Figure 1 describes the stages of assessment development in very broad terms. There are, of course, many specific substeps within each stage. These substeps need to be specified in the local context before the model can be implemented. The substeps may vary depending on the curriculum and the country. Figure 4 shows the substeps in an adaptation of the model with which Botswana's Department of Curriculum Development and Evaluation is experimenting. This department is trialing the model with end-of-term tests, with plans to implement it with the primary school leaving examination.

Insert Figure 4 here

Some Expected Benefits of Curriculum-Driven Criterion-Referenced Examinations

When examinations, curriculum, and classroom teaching are linked together in the seamless fashion described in this paper, we can expect some important educational benefits. I list these benefits below.

1. **Improved curriculum implementation** Feedback to schools and teachers that focus on how students performed in specific curriculum areas or on clusters of curriculum objectives reinforces teachers for teaching the official curriculum.

2. **Fairness to students** If the curriculum is clearly defined, if the examination plan is known and understood by teachers, if teachers teach toward this examination plan, and if the examinations mirror the plan and curriculum, then the examinations become fair to students because they will have been taught what is expected of them on the examination. Fairness is closely linked to the principal of seamlessness described previously.

3. **National educational progress can be evaluated** Curriculum-driven criterion-referenced examinations permit one to analyze the results of the examination to describe what the nation's students are capable doing. Since the examination specifications (assessment plan) remains constant over several years, one may monitor progress on specific curriculum learning targets by comparing, over the years, the percentage of the nation that has learned each target. Students' performance on clusters of curriculum objectives (e.g., those dealing with solving nonroutine problems) can be compared as well.

4. **Improved curriculum evaluation** One aspect of curriculum evaluation is the extent to which each curriculum learning target is learned. Data from curriculum-driven criterion-referenced tests may be used to identify which learning targets have been learned better than others.

5. **Career guidance for individuals** One part of career guidance consists of learning one's strengths and weaknesses, one's skills and abilities. Criterion-referenced interpretations contribute to this knowledge because they describe the degree to which each part of each curriculum has been mastered. This provides a rather specific profile of a student's knowledge and skills that may be used for guidance purposes. (Much more than this is needed for proper guidance, of course.)

6. **Better diagnosis of student's deficiencies** If periodic and grade-level curriculum-driven criterion-referenced assessments are created, then teachers could receive information about the degree to which each student has learned specific learning targets. Knowing which learning targets have not been learned sufficiently can help a teacher focus remedial instruction where it is needed.

7. **More focused target inservice teacher training** Education field officers (inspectors) can be provided information about the performance of students on specific areas of each curriculum for each school they monitor. If patterns emerge over time to indicate specific areas of the curriculum not being taught well, inservice teacher training can be effectively targeted on a school-by-school basis.

8. **Continuous assessment possibilities** Since the curriculum is the basis for building examinations, it is also the basis for developing other assessments to monitor students' progress toward important learning targets. For example, curriculum-driven criterion-referenced assessments could be developed for the learning targets that we expect students to learn each term. Teachers could administer these termly assessments and use the results to identify each student's progress and to provide remediation where possible. Teachers determined easily which students are "on-target" and which are appearing to fall behind. Termly assessments that focus on those curriculum learning targets that were ***actually taught to students*** should be used to assign a grade to each student. If each school's assessments are based on the same curriculum learning targets, student grades may have a more consistent and meaningful foundation. (A curriculum-driven criterion-referenced continuous assessment framework is described in detail in another paper [Nitko, 1994a]).

Summary

In this paper I have proposed a way of developing national examinations in the context of curriculum reforms and new programs for expanding basic education. I proposed a model that creates a strong alignment of national examinations and national basic education curricula.

In a high-stakes environment, where examinations determine who is certified and selected for further education, examination development cannot proceed independently from national curriculum reform. It is necessary for persons at all levels of the educational enterprise to understand that teaching all important elements of the new curricula is the best way to prepare students for the national examinations. Articulating curricula and examinations requires at least three components: (1) a formal policy statement from the ministry stating the necessity for this articulation, (2) adopting a curriculum-driven criterion-referenced examination development model that gives the specific steps required for developing examinations, and (3) establishing an oversight committee to ensure that the required policy and development model are implemented to the satisfaction of the Ministry of Education.

The curriculum-driven examination development model I presented is different than a traditional "syllabus" driven model. A curriculum-driven model requires that the official national curricula be the center of the examination development process. Decisions about what and how to examine are heavily influenced by the curriculas' stated learning outcomes. A curriculum-driven approach also implies quite a different role for each subject area's "examination committee" and for the "chief examiner". Their roles become more facilitative than authoritative because the curriculum defines the content and performance levels to be examined.

The curriculum-driven model I discussed requires more empirical research and data analysis than has traditionally been done regarding leaving examinations. The model proposes

going beyond trialing examination questions and simple reporting of results. It proposes formal adoption of educationally and psychometrically sound quality control standards. Empirical research data are used to measure and monitor examination quality against the specified standards. The model also proposes conducting secondary analyses of the examination results, and studying the performance of students in relation to the various components of the curriculum. The purpose of these analyses is to provide outcomes-based information for curriculum improvement, for national educational monitoring and policy formation, and for monitoring individual schools so that curriculum-based inservice programs may be targeted to them.

Among the benefits expected from implementing curriculum-driven examinations are improvements in (1) implementing curriculum reforms, (2) examination fairness, (3) assessment of national educational progress, (4) outcomes-based curriculum evaluations, (5) career and job guidance, (6) teacher attention to curriculum areas needing instruction, (7) targeting of inservice teacher trained specific to schools, and (8) articulation of continuous assessment with the national examinations.

References

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L.Thorndike (Ed.), ***Educational measurement*** (2nd ed.). Washington, DC: American Council on Education.

Holland, P.W., & Rubin, D.B. (Eds.) (1982). ***Test equating***. New York: Academic Press.

Jaeger, R.M. (1989). Certification of student competence. In R.L. Linn (Ed.), ***Educational measurement*** (3rd ed.). New York: Macmillan, 485-514.

Lindquist, E.F. (1951). Preliminary considerations in objective test construction. In E.F.Lindquist (Ed.), ***Educational measurement***. Washington, DC: American Council on Education, 119-184.

Livingston, S.A. & Zieky, M.J. (1982). ***Passing scores: A manual for setting standards of performance on educational and occupational tests***. Princeton, NJ: Educational Testing Service.

Nitko, A.J. (1989). ***Criterion-referencing junior secondary school student assessments***. Gaborone, Botswana: Junior Secondary Education Improvement Project, Learning Systems Institute, Florida State University.

Nitko, A.J. (1994a). ***Curriculum-based criterion-referenced continuous assessment: A framework for the concepts and procedures of using continuous assessments for formative and summative evaluation of student learning***. A paper presented at the Second International Conference on Educational Evaluation and Assessment, Association for the Study of Educational Evaluation in Southern Africa, Pretoria, July.

Nitko, A. J. (1994b). ***Is curriculum a reasonable basis for assessment reform***? Presidential address at the annual meeting of the National Council on Measurement in Education, New Orleans, April.

Peterson, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming, and equating. In R.L. Linn (Ed.), ***Educational measurement*** (3rd ed.). New York: Macmillan, 221-262.

Popham, J. W. (1992). ***A tale of two test-specification stratagies. Educational Measurement: Issues and Practice***, **11**(2), 16-17,22.

Posner, A. (1992). ***Analyzing the curriculum***. New York: McGraw-Hill.

Table 1. Comparison of roles and Functions of the Examination Committee Under Traditional Operations and Under Curriculum-Drive Criterion Referenced Examination Development Examination.

Traditional Examination Development Operations	Curriculum-Driven Criterion-Referenced Examination Development
1. Chief examiner	1. (a) No chief examiner role (b) Examination Committee Coordinator coordinates and facilitates the development of the examination
2. Chief examiner and/or committee member set examination syllabus and test plan.	2. (a) No examination syllabus (b) Team to develop examination specifications and plan
3. Chief examiner and/or committee members may set examination questions.	3. (a) Committee members set no questions (b) Item specifications developed by curriculum examination officers. (c) Committee reviews/moderates item specifications in light of the examination specifications and plan.
4. Chief examiner and/or committee members review and select questions comprising the examination. Questions may be moderated	4. (a) Questions set using item specifications by teachers, edited by curriculum/ examination team members, tried-out by examination officers, and revised by curriculum/examination team to conform item specifications. (b) Using the examination specifications and assessment plan team members develop draft of examination to present to the Examination Committee. (c) Examination Committee reviews the proposed examination and approves/moderates it.
5. Chief examiner and/or committee members set marking schemes and/or answer keys for examination papers.	5. (a) Marking schemes and/or answer keys set by team members. (b) Examination committee reviews the proposed marking schemes and/or answer keys and approves/moderates them.
6. Chief examiner and/or committee members supervise the marking of examinations where necessary. Marks may be moderated.	6. Examinations committee coordinator and/or committee members supervise the marking of examinations where necessary.
7. (a) Examination committee sets grade boundaries based on examinees' performance, existing policy, and weighing of examination and continuous assessment components. (b) Examination unit analyzes the results, summarizes the score distribution in relation to the grade boundaries and past years' results, and presents analyses to Examination Committee at "awards meeting". (c) Examination Committee reviews grade boundaries and may adjust them after review of data in order to maintain comparable quality standards from year to year.	7. (a) Curriculum/examination officer team propose percentage weighing of the examination and continuous assessment components. (b) Examination committee reviews, approves and/or moderates the recommended grade boundaries. (c) Examination unit analyzes examination results item-by-item in relation to the specification, summarizes the score distribution in relation to the grade boundaries and past years' results, and presents analyses to Examination Committee at an "awards meeting". (d) Examination Committee reviews grade boundaries and may adjust boundaries after review of data in order to maintain comparable quality standards from year to year.

Table 2. Examples of Quality Control Ares, Standards, Measures, and Criteria for Curriculum-Driven Criterion-Referenced National Examination

Quality Control Area	Standard	Measure	Criterion to be Met By Each Item
Quality of the content of the test items	1. Accuracy of the content	1. Contents experts' ratings of each item (0-4).	1. Average rating of 3.5 per item.
	2. Accuracy of the keyed answer.	2. Content experts' judgements (yes, no).	2. All experts agree that the keyed answer is correct or the best choice.
	3. Relevance and importance of the task to be performed.	3. Content experts' ratings of each item (0-4).	3. Average rating of 3.5 for each item.
	4. Congruence of test item to the objective.	4. Ratings of knowledgeable teachers (0-4).	4. Average rating of 3.5 per item.
	5. Correspondence of test item to thinking skill category.	5. Ratings of knowledgeable teachers (0-4)	5. Average rating of 3.5 per item.
Technical quality of individual test items	1. Flawlessly written items.	1. Review of item by professional item-writer.	1. Each item must not exhibit any item writing flaw.
	2. Appropriate vocabulary.	2. All words in the item are from the designated vocabulary list(s).	2. Each item contains only those word on the designated list(s).
	3. Appropriate difficulty.	3. Item p-value from tryout sample.	3. .05<p<.95
	4. Appropriate difficulty.	4. Item discrimination Index.	4. r>.2
	5. Avoidance of ethnic and gender stereotypes.	5. Judgments of representatives of affected groups.	5 No item judged to contain a stereotype.
	6. Avoidance of bias and offensiveness.	6. Judgments of representatives of affected groups	6. No item judged to contain a stereotype.
Quality of the test scores	1. Same distribution of item difficulty indexes on every year's test for a subject (distribution may vary for different subjects).	1. Compare the actual item difficulty distribution against the specified distribution.	1. Every test must meet the specified distribution before it is used.
	2. Same distribution of item discrimination indexes on every year's test for a subject (distribution may vary for different subjects).	2. Compare the actual item discrimination distribution against the specified distribution.	2. Every test must meet the specified distribution before it is used.
	3. High reliability.	3. Coefficient alpha or Kuder-Richardson 20	3. Each test should have a coefficient greater than or equal to .85.
	4. High marker reliability for essays and practicals.	4. Percent agreement.	4. Each test should have a percent agreement of .90 or higher.
	5. High decision consistency.	5. Kappa coefficient.	5. Each test should have a kappa value of .6 or higher at the designated passing scores.
	6. High convergent validity.	6. Correlation coefficients.	6. r>.60 between test scores and continuous assessment grades in immediate past and at the next level (After correction for restriction in range).

Table 3. Examples of Stakeholders for receiving reports of the result of curriculum-driven criterion-referenced national examinations.

Stakeholder	Example Type of information to be reported
Principal secretary of education	1. Distribution of school averages in each subject for the nation.
	2. Distribution of school averages in each subject for each region (state)
	3. Trend of region (state) and national averages in each curriculum subject over the past five or ten years.
	4. Graph that simultaneously compares curriculum achievement outputs to educational inputs (e.g., teacher qualifications, school resources) over five or ten years for each region (state) and nation.
Curriculum development officer	1. Average performance of students on clusters of questions assessing each curriculum learning target.
	2. Average performance of students in broad topical areas within a curriculum.
	3. Results in 1 and 2 above reported by nation, by region, by state.
Educational field officer (inspector)	1. Average results of individual schools within the service region in each curriculum area.
	2. Average results by broad topical areas within a curriculum subject of individual schools within the service region.
	3. Trend of each school's average performance over the past five years.
Headmaster (Principal)	1. Average results of students in the particular school showing topical areas within each curriculum.
	2. Same as 1 except including a comparison of the particular school with the national distribution of school means.

Table 4 Example of a curriculum-driven school report

NAP '94 Grade 6 Math Test Results for		[School name has been removed from this copy]					
Student Name	Math Area:	Number (Percent)	Geometry (Percent)	Measurement (Percent)	Statistics (Percent)	Algebra (Percent)	Total (Percent)
		absent					
		71	63	54	77	99	70
		67	27	45	22	99	55
		45	18	27	33	33	37
		32	18	27	55	0	31
		43	9	27	22	66	35
		32	27	18	22	66	30
		absent					
		82	63	81	77	66	78
		absent					
		65	27	45	77	99	60
		absent					
		absent					
[Student		47	9	18	33	33	36
names have		32	18	54	33	0	32
been removed		78	54	72	88	66	75
from this		49	27	27	11	0	37
copy]		60	27	81	77	33	60
		34	0	9	44	0	26
		17	18	18	11	0	16
		93	72	81	77	99	87
		91	63	99	77	99	87
		absent					
		absent					
		91	72	72	77	99	85
		32	27	18	22	33	28
		absent					
		34	9	45	33	33	32
		23	9	9	33	0	20
		47	18	72	55	66	48
		95	63	81	77	66	86
		absent					
		absent					
		28	27	18	22	33	26
		82	54	54	55	66	71
		93	72	72	77	99	86
		absent					
		absent					
		36	27	9	11	33	28
		17	0	27	11	66	17
		36	36	18	11	33	31
		39	18	9	33	66	32
		86	72	72	77	99	82
School average - males		57	37	50	52	58	53
School average - females		52	31	39	41	49	46
School average - overall		54	34	44	46	53	49
School vs nation comparison		High	High	High	High	High	High
National averages		43	31	40	32	43	40

Percentage absentees = 28 and with poor marking/unscorable papers = 0

Source: Adapted from the Ministry of Educationand Culture, Jamaica

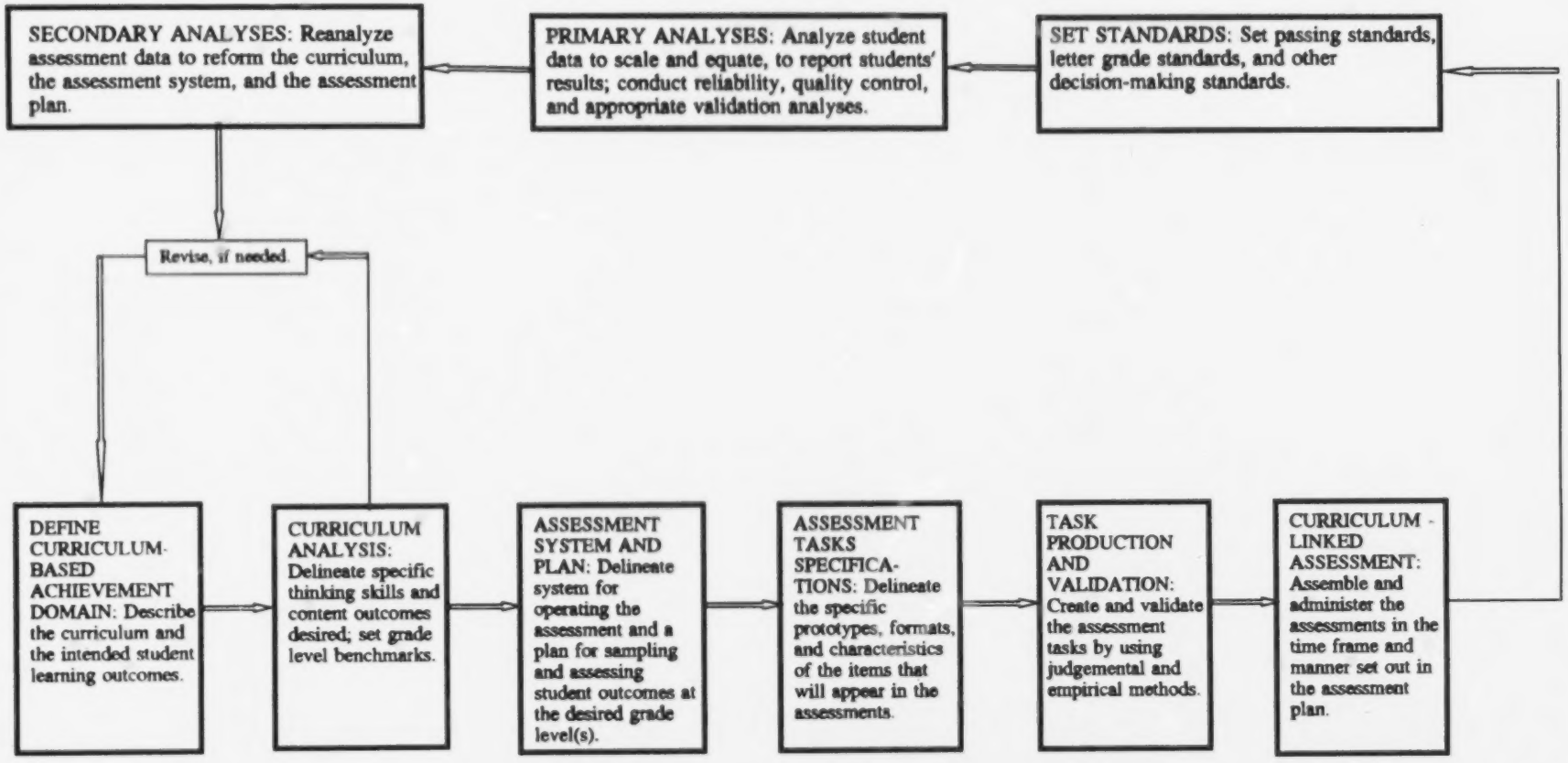
SECONDARY ANALYSES: Reanalyze assessment data to reform the curriculum, the assessment system, and the assessment plan.
PRIMARY ANALYSES: Analyze student data to scale and equate, to report students' results; conduct reliability, quality control, and appropriate validation analyses.
SET STANDARDS: Set passing standards, letter grade standards, and other decision-making standards.
Revise, if needed.
DEFINE CURRICULUM-BASED ACHIEVEMENT DOMAIN: Describe the curriculum and the intended student learning outcomes.
CURRICULUM ANALYSIS: Delineate specific thinking skills and content outcomes desired; set grade level benchmarks.
ASSESSMENT SYSTEM AND PLAN: Delineate system for operating the assessment and a plan for sampling and assessing student outcomes at the desired grade level(s).
ASSESSMENT TASKS SPECIFICATIONS: Delineate the specific prototypes, formats, and characteristics of the items that will appear in the assessments.
TASK PRODUCTION AND VALIDATION: Create and validate the assessment tasks by using judgemental and empirical methods.
CURRICULUM - LINKED ASSESSMENT: Assemble and administer the assessments in the time frame and manner set out in the assessment plan.

Figure 1. Process Model for Curriculum-Driven Assessment Development

Performance Indicator (Objective)

Use criteria for determining particularly misleading ads to identify such ads.

Rationale

The realities of inflationary prices and the declining quality of many manufactured products mandate closer scrutiny of advertising by consumers in order to protect themselves and their investments. The ability to recognize advertising which misrepresents a product or service is crucial for the individual and for the general welfare of the country. Individuals who can identify misleading ads will be in a position to purchase better products, allowing them to save money over the long run.

General Description

The student will be presented with four or five product or service advertisements. Multiple-choice questions will be designed to determine if the student can use criteria for determining misleading ads to identify misleading information, poor advertising practices, and/or misrepresentation of the product or service advertised. The ads will be presented in their original form, as in actual ads, or be specifically written for the test.

Sample Item

[Presentation of four ads (not reproduced in this illustration)]

Which of these ads can be considered misleading because it uses excessive language to sell the product?

*A. the ad for body building
B. the ad for astringent cleanser
C. the ad for ice cream
D. the ad for toothpaste

Stimulus Attributes

The general stimulus for this performance indicator should contain four or five sample ads, developed or selected and presented according to the following guidelines:

1. For each item (or set of items), four or five actual or simulated advertisements drawn from a variety of sources (newspapers, magazines, radio, television, etc.) should be presented.
2. One set of ads may be used for several items.
3. Ads should describe products or services designed for only males or for only females in addition to products or services designed for both sexes.
4. Ads should describe products or services designed for the general age level of the students being tested.
5. Ads may be of three types:
   a. written
   b. written with illustrations or pictures
   c. oral (presented on a tape recorder)
6. Ads should be no longer than one typewritten page or 300 words.
7. A minimum of four and a maximum of five ads should be presented for any one item or set of items.

Stem Attributes

1. Following the presentation of the ads, there should be either a single item or a set of items.
2. Item stems should ask students to identify which ad is misleading according to a specified criterion.
3. The criteria should come from the following list. An ad may be misleading if it:
   a. creates an impression that is different from the single statements or facts presented, even though every statement is correct.
   b. conceals important facts about the product or service (e.g., price, guarantees).
   c. diverts attention from the actual terms and conditions of the offer.
   d. makes false or misleading comparisons with other products or services.
   e. makes an offer that appears to be too good to be true, thus creating false expectations.
   f. appeals to ideas or sentiments that are loved, cherished, or respected by many people (e.g., the family or patriotism), otherwise known as "flag waving."
   g. appeals to scientific authority or documentation.
   h. appeals to one's desire to be part of the group, up with the times, in tune with the latest fad, otherwise known as the "bandwagon approach."
   i. employs "snob appeal" by using famous individuals or people from prestigious groups or occupations to advertise the product or service.
   j. uses many superlatives and other forms of excessive language (e.g., the best, the newest, the greatest) to try to sell the product or service, otherwise known as "glittering generalities."

   Only criteria listed above may be used.
4. The stem should be written in language not to exceed the seventh-grade reading level.

Response Attributes

1. The responses should follow a four-alternative multiple choice format.
2. The correct response should be the name or a brief description of the only ad that is misleading for the reason given.
3. Distractors should be the names or brief descriptions of the ads that are either not misleading or are misleading for reasons other than the one given.

Source: Adapted from the Rhode Island Statewide Assessment Program, 1980.

Figure 2. Example of a detailed item-specification for curriculum-driven examination

Mid-Level Test-Item Specifications

Items may call for students to create or choose the most accurate summary of the selection or part of the selection, to identify or state the topic of all or a part of the selection, or to identify or state the main idea or central point of a selection or part of that selection. Students may have to condense explicit information, or to paraphrase or restate points, but should not have to make an inference in order to select or construct the appropriate answer. Items can be phrased in a variety of ways, but they all *must require the student to have recognized the central message or overall point of the selection* (or designated part of the selection).

Sample Items

What is this selection mainly about?
Write a brief paragraph summarizing this passage.
Which of these options BEST summarizes the article?
Describe, in one sentence, the passage's central message.
What is the main point of this passage?
What is the main idea of the passage's fourth paragraph?

Source: Popham (1992)

Figure 3. Example of a mid-level item-specification for curriculum-driven examination

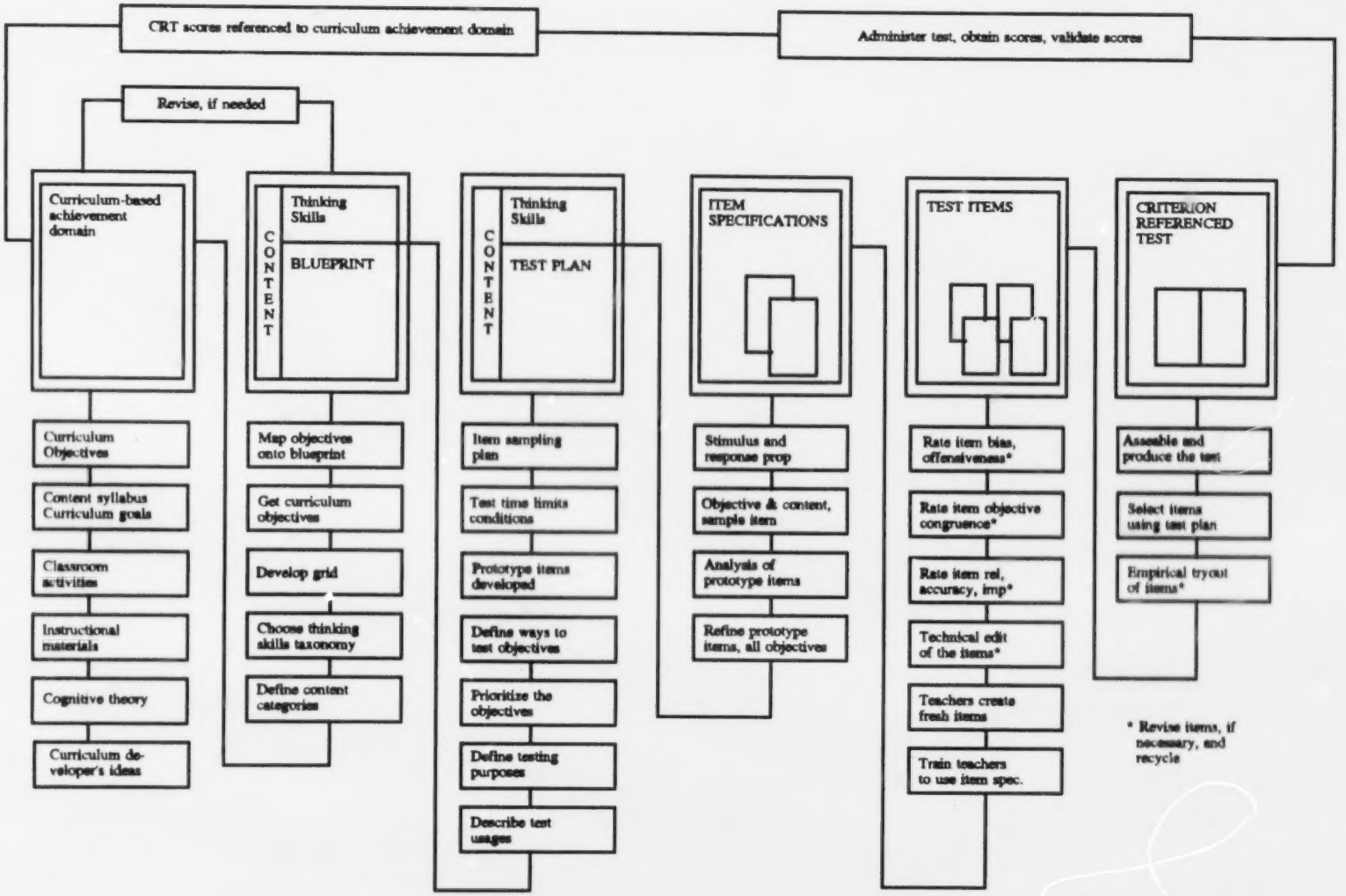

Figure 4. Fleshed-Out Details for Curriculum-Driven Criterion-Referenced Test Development In Bots3wana

BEST COPY AVAILABLE